ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Energy Procedia 37 (2013) 4057 – 4064

Energy Procedia

GHGT-11

Application of an unsupervised methodology for the indirect detection of CO_2 leakages around the Laacher See in Germany using remote sensing data

Rajesh Govindan, Anna Korre*, Sevket Durucan

Department of Earth Science and Engineering, Royal School of Mines, Imperial College London, London SW7 2BP, UK

Abstract

Remote sensing has demonstrated success in various environmental applications and is considered attractive as it offers cheap and effective wide area monitoring. Recent research efforts, and particularly the development of an unsupervised image processing and indirect detection methodology by the authors, have demonstrated the potential of remote sensing techniques for monitoring CO_2 storage sites. This paper describes the application of the methodology at a natural analogue study site around the lake Laacher in Germany using airborne optical remote sensing datasets. The results demonstrate the robustness of the methodology as an indirect technique for CO_2 leakage detection.

Keywords: CO_2 storage; remote sensing; Independent Component Analysis; Self-Organising Maps; Dempster-Shafer theory

1. Introduction

Climate change is a problem of global concern that poses many challenges for scientists today. The presence of ever increasing quantities of CO_2 gas in the atmosphere emitted from fossil fuel utilisation contributes significantly to this problem. Recent technological research and development has focused on investigating the concept of Carbon dioxide Capture and Storage (CCS), wherein CO_2 is separated from flue gases, transported and stored in deep geological formations. In the wake of the introduction of CCS technology, issues from science and engineering to economics are investigated by researchers to ensure that CCS provides a safe and sustainable option in the climate change mitigation solutions portfolio.

* Corresponding author. Tel.: +44-20-7594-7372; fax: +44-20-7594-7444.
E-mail address: a.korre@imperial.ac.uk

doi:10.1016/j.egypro.2013.06.306

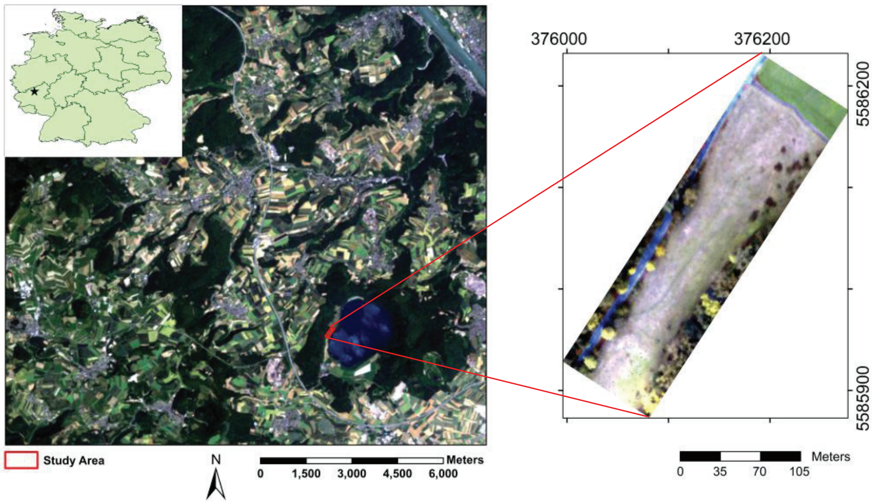
However, it is important to note that simple physical factors such as the pressure gradient between the surface and the reservoir, and the existence of migration pathways, such as faults and abandoned wells, could potentially cause stored CO_2 to migrate towards the surface, eventually resulting in leakage. This potential for leakage, and that it could ultimately lead to a regional ecological imbalance, is the source of significant public concern. For this reason suitable technologies that can reliably monitor and detect potential CO_2 leakages on the surface are of great value. Today, airborne and spaceborne platforms for optical remote sensing of the Earth are playing an important role in environmental monitoring applications. This technology is attractive as it offers a wide range of sensors that can simultaneously record images in hundreds of contiguous regions of the electromagnetic spectrum. More importantly, optical remote sensing can also provide good areal coverage, which is attractive considering that a large surface area may be affected by potential CO_2 seepages from a leaking storage complex.

Existing methods that are used to monitor CO_2 leakage at the surface are divided into two general groups: direct methods and indirect methods. The direct methods involve field-based measurement techniques to detect and quantify gas leakages, while indirect methods involve studying the effects of leakage on the surface environment, such as vegetation stress patterns and mineral alterations in the soil, that are further validated using data acquired through direct methods. However, the fundamental problem associated with existing methods is that they involve large amount of manual effort in data collection and visual interpretation that could be error prone or time consuming. It is therefore desirable to enhance these methods or develop new ones that can first detect locations of potential CO_2 leakages in an unsupervised manner. Once such leakage locations with high confidence are known, it is possible to design field surveys to minimise the area covered and time spent for collecting direct measurements.

This paper builds upon the development of an indirect and unsupervised methodology presented by the authors before [1], using airborne optical remote sensing imagery acquired over Maria Laach, Germany, where natural seepage of CO_2 occurs on the surface. The underlying idea behind the proposed methodology is to model the spatial correlations in the images studied and filter out data redundancies using the geostatistical theory of Intrinsic Random Functions of order-*k* (IRF-*k*). The aim is to obtain residual signals, and is followed by the estimation of a hidden residual variable using Independent Component Analysis (ICA). The hidden variable, as a prior distribution, is then combined with a fuzzy evidence (the likelihood distribution), which conveys the surface effects due to CO_2 leakage, using the Growing Hierarchical Self-Organising Maps (GHSOM) and the Dempster-Shafer (DS) theory of evidence combination. The result is a posterior confidence map showing anomalous areas related to leakage on the surface. The posterior result is also validated using available ground measurement data.

2. Study area and data available

The study area is located near a lake in Maria Laach, a nearly circular flooded caldera in the east Eifel volcanic district of the Rhenish Massif in Southwest Germany. This is the largest maar lake, compared to those in the west Eifel region, which formed as a result of the most recent eruption of a huge volcano producing some 5 km^3 of highly evolved mafic phonolite magma, about 11,000 years BP [2, 3]. Today, the lake in the Maria Laach caldera has a maximum depth of 52 m and a surface area of 3.31 km^2, covering 27 % of its drainage area. Degassing of up to 99 % vol. of CO_2 occurs from the mantle, as indicated by the Helium and Carbon isotope data collected in the area [4]. Surface gas emissions have been found from several points in and around the lake [3]. Release to the atmosphere typically occurs from gas vents, characterised by a small core of elevated gas flux. Discharge of CO_2 along the eastern shore of Laacher See has been known for many years. Rising gas bubbles can be observed on the water surface at some locations near the Eastern shore. The field survey conducted in 2008 by Jones et al. [5] reported occurrence of CO_2 in the western shores of the lake with concentrations of up to 1,600 ppm



(a) (b)

Fig. 1. (a) The region of Laacher See in Germany; (b) the location of the study area on its western shore.

against an average background level of around 350 ppm.

The area of interest chosen for the study presented here is a small section of the terrestrial surface along the western shores of the lake, as illustrated in Fig. 1. The optical remote sensing data for the study area is airborne hyperspectral imagery, acquired by the Instituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Trieste, Italy, through surveys conducted in 2009, using the AISA Eagle 1K sensor at 1m resolution and containing 63 bands, ranging from visible through Near Infra-Red (NIR) wavelengths (0.40 - 0.99 μm). The mobile laser measurement survey results for CO_2 concentrations (in ppm) conducted in 2008 [5] by the British Geological Survey (BGS) are used for the validation of the results.

3. Methodology

The methodology developed by the authors has already been tested using optical remote sensing datasets for the Latera natural CO_2 emission site in Italy [1, 6]. The main principles of the methodology are discussed in this section.

Given the volume of information contained in the hyperspectral imagery, it is essential to perform pre-processing steps that can reduce factors such as data corruption by noise and high dimensionality before dealing with specialised problems such as the detection of CO_2 leakages. Hence, prior to the application of the methodology, data pre-processing using data noise reduction by Maximum Noise Fraction (MNF) and dimensionality reduction by Principal Component Analysis (PCA) was applied. Using the MNF transformation, the noisy vectors in the data were first discarded. Once the noise was sufficiently minimised, the dimensionality of the data was reduced (reducing the number of bands) using a band grouping scheme based on PCA. The pre-processed data was then analysed using the methodology discussed below.

3.1. Geostatistical filtering for prior detection

The pre-processed hyperspectral imagery was considered as a 2D Random Function (RF), Z(**p**), which is linearly represented by the sum of a deterministic part referred to as the mean (trend), and a stochastic part referred to as the non-stationary spatial variability. This variability is represented by an error model, such as the covariance function, which characterises the heterogeneity in radiometric values of image pixels (or image texture). Kitanidis [7] proved that the variogram function, which is based on the intrinsic hypothesis of random and structured spatial variables, approximates the covariance function. The classical definition of the variogram function is given by:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})}\sum_{i=1}^{N(\mathbf{h})}\left[Z(\mathbf{p}_i) - Z(\mathbf{p}_i + \mathbf{h})\right]^2 \tag{1}$$

where, $\gamma(\mathbf{h})$ is the variogram function, $Z(\mathbf{p}_i)$ and $Z(\mathbf{p}_i + \mathbf{h})$ are pairs of values (realisations) of the random variable at locations $\mathbf{p}_i$ and $\mathbf{p}_i + \mathbf{h}$ separated by the lag (displacement) vector $\mathbf{h}$, and $N(\mathbf{h})$ is the number of pairs available for a given lag. Hence, the variogram function serves as a spatial structure analysis tool, and unlike the covariance function it does not require the estimation of the usually unknown mean value of Z, thereby simplifying parameter estimation. However, in order to avoid bias in estimation due to the presence of spatial heterogeneity, the theory of IRF-*k* [8] was applied to the hyperspectral imagery to obtain stationary residuals. Independent Component Analysis (ICA) [9] was then applied to linearly estimate hidden non-Gaussian variables from these residuals for the prior detection result. Non-Gaussian behaviour is assumed based on the fact that the presence of potential CO_2 leakage anomalies (outlier pixels) in the images causes the data distributions to become positively skewed.

3.2. Unsupervised learning and fuzzy image clustering

Soft (or fuzzy) classification was used to generate the leakage likelihood distribution since it has the advantage of relaxing the crisp membership condition imposed by the conventional approach of hard classification, and allowing a given pixel to have partial membership under all the predefined pixel classes, particularly for the case of low resolution data where the problem of mixed pixels is plausible. The likelihood distribution that was thus generated is based on classes corresponding to the surface effects caused by CO_2 leakage, namely vegetation stress and soil mineral alteration. This was done using unsupervised learning and fuzzy clustering of the hyperspectral imagery (not pre-processed for dimensionality reduction) based on Self-Organising Maps (SOM) [10]. The fundamental idea is to map a high dimensional dataset's feature space onto a 2D representation space, by means of a *competitive learning* neural network model. Competitive learning is the non-linear mapping of n input data vectors $\mathbf{z}_i$ ($i = 1, \dots, n$) from an N-dimensional domain R^N to a 2D map of m N-dimensional output weight vectors $\mathbf{w}_j$ ($j = 1, \dots, m$; $m < n$), ensuring that the original topology of the data is maintained, i.e., similar pixels are grouped close to each other in the new space. In the classical theory of Vector Quantisation (VQ), this set of output vectors is called the *codebook*. The change in the state of the codebook (adaptation) is dependent on input such that the codebook vector closest to a given $\mathbf{z}_i$, called the Best Matching Unit (BMU) or the 'winner' $\mathbf{w}_j$, is always updated. This is done using the *steepest-descent gradient step optimisation*, given by:

$$\mathbf{w}_C(t+1) = \mathbf{w}_C(t) + \alpha(t)\left[\mathbf{z}_i(t) - \mathbf{w}_C(t)\right] \qquad \mathbf{w}_j(t+1) = \mathbf{w}_j(t),\ j \neq C \tag{2}$$

where, $j = C$ is the BMU (or winner) index, t is the time (iteration) index, and α is the adaptation gain coefficient, whose value lies between 0 and 1. However, owing to SOM's fixed architecture (number of neurons) and its inability to maintain the inherent hierarchical structure of data, a dynamic version called GHSOM [11] was used instead. Moreover, GHSOM enables the automatic discovery of all natural clusters in a multivariate dataset, avoiding the need for the user to randomly guess the number of classes, which is usually a hard decision to make for data with a wide spatial coverage. This also improves the likelihood distribution and potentially lowers the rate of false detections.

3.3. Probabilistic image fusion for posterior detection

In the Bayesian context, the methodology aims at improving the *prior detection map*, obtained after using the geostatistical filter analysis and ICA, by incorporating information provided by the *fuzzy clustering map* regarding the likelihood of CO_2 leakages on the surface. This generates a confidence map as the posterior detection of leakages, based on an information fusion theory called the Dempster-Shafer (DS) theory of evidence combination [12, 13]. The DS theory is a generalised probabilistic model where probability values are based on sets of joint hypotheses rather than just mutually exclusive or singleton sets. If the hypothesis space contains n singletons, then a *frame of discernment* is formed by a power set θ containing 2^n elements, or possibilities. The null set ϕ is one of the elements of the power set that represents a scenario of conflicting evidence between two or more hypotheses in the system. Hence, discounting the null set, the actual sum of all probabilities in the frame of discernment can be less than or equal to unity, unlike classical probability. Like the popular Bayesian inference, the DS theory combines hypotheses (evidences) as and when available to update the system's prior belief.

Information fusion was performed by representing each of the aforementioned maps as Probability Mass Functions (PMF) and applying the DS combination rule in order to estimate the Basic Probability Assignment (BPA) per image pixel, with an uncertainty range [*belief*, *plausibility*]. The rule of evidence combination of multiple PMF functions is a generalisation of the Bayes' rule, called the DS combination. It strongly emphasises the agreement between multiple sources and eliminates conflicts through a normalisation factor. This can also be considered as a strict AND-operation in fuzzy logic terms. For combining a pair of PMF functions of independent evidential sets, say X and Y, in order to generate a new composite hypothesis set, say A, DS combination rule states that:

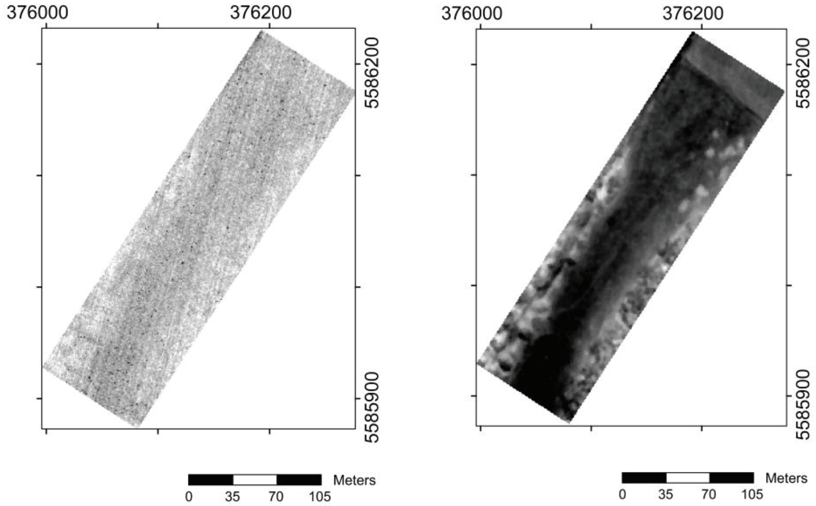
$$K = \sum_{X \cap Y = \phi} m(X)\, m(Y) \qquad m_X \oplus m_Y(A) = \frac{\sum_{X \cap Y = A} m(X)\, m(Y)}{1 - K} \tag{3}$$

where, m() is the PMF, and 1 - K is the normalising factor, which has the effect of removing conflicts and attributing any PMF associated with it to the null set. When K = 0, the belief and plausibility values are equal, as in the case of classical probability. Finally, those pixels with low BPA values and low uncertainty are converted as high confidence measures using the entropy function from information theory, indicating leakage anomalies as posterior detection map [1, 6].

4. Results and discussion

4.1. Pre-processing results

As shown in Fig. 2a, the hyperspectral image bands were severely corrupted by noise particularly for the NIR region. This noise could have been exaggerated by the geometric correction that was performed prior to making the data available for this study. Geometric transformation into a 2D coordinate system, such as the Universal Transverse Mercator (UTM) coordinates in this case, smears pre-existing acquisition noise in the data thereby decreasing its initial quality. Hence, it is recommended that geometric transformation should always be done only after noise in the data has been significantly reduced using the MNF transformation. Fig. 2 illustrates the result obtained for Band 1 (0.402 μm) after applying MNF on the Laacher See data.



(a) (b)

Fig. 2. (a) Band 1 of the 2009 hyperspectral data for Laacher See corrupted by noise; (b) Band 1 after applying MNF.

4.2. Prior and Posterior detection results

The pre-processed multivariate data was subsequently filtered by using the IRF-*k* theory, implemented in the geostatistical software package Isatis. The co-kriging system uses spatial correlation information from the modelled variogram function to estimate the weights of the filter. The filtered data (or residuals) was then used to estimate the prior detection through ICA. Hence, a pool of non-Gaussian Independent Components (ICs) are obtained, only one of which best represents the prior anomaly distribution. It was then assumed that in order to represent a prior distribution, one of the candidates from the pool must maintain spatial homogeneity in variance (or homoscedasticity). This property was considered essential for selection because it implicitly conveys the fact that all the pixels in the IC are identically distributed, i.e. their prior probabilities as anomalous pixels are equally likely [1, 6]. Fig. 3a is the prior detection map obtained after selection based on the variance criterion. Unsupervised learning and clustering was then applied to the dataset using the GHSOM. Two codebook vectors (or spectral signatures) were identified to be significant, representing: (a) the class of stressed vegetation; and (b) the class of healthy vegetation. Using the signature corresponding to the former class, fuzzy analysis was performed to obtain the likelihood map for potential CO_2 leakage, as shown in Fig. 3b. It is shown that the stressed vegetation areas are concentrated in the region close to the shore of the Laacher See. Furthermore, the DS theory was applied per-pixel on the prior detection and likelihood maps obtained, after estimating their respective PMFs. The posterior detection map obtained by information fusion, in Fig. 3c, shows high confidence in the regions close to the lake shore.

The posterior detection result was further validated using ground truth data acquired using a mobile laser device [5]. The point measurements of CO_2 concentration were used to estimate the mean spatial distribution of CO_2 concentration using the block kriging technique (Fig. 4a). The corresponding standard deviation of the kriged estimation was then used to map ground truth data locations lying beyond two standard deviations from the mean estimates. Using this map as a filter, locations of anomalous

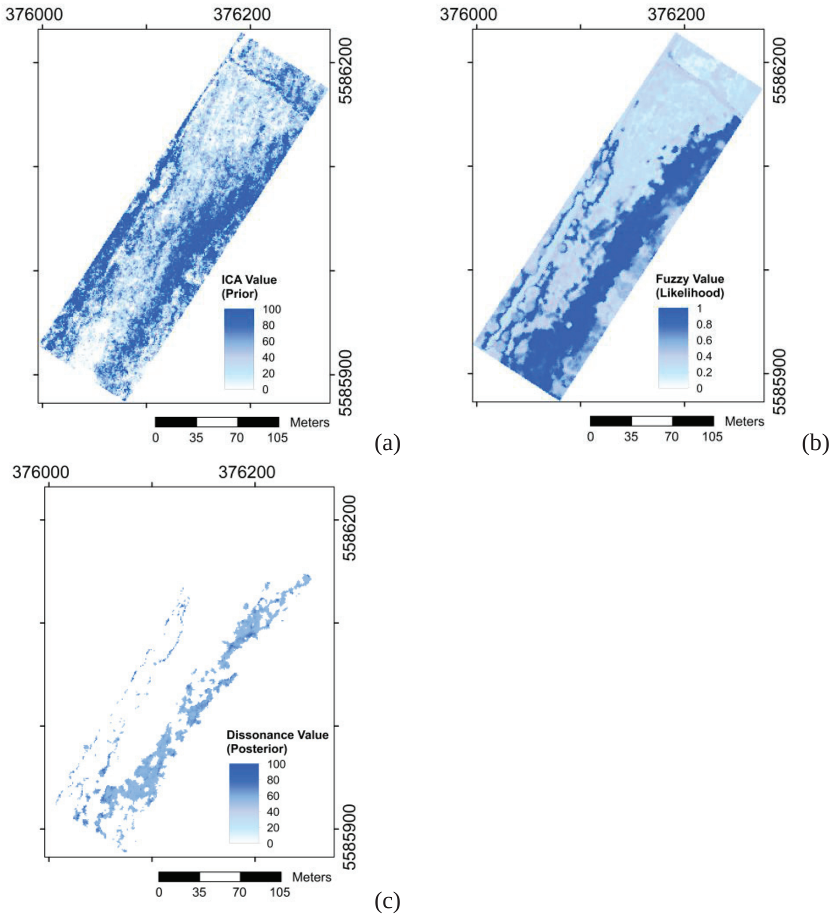


Fig. 3. (a) Laacher See data processing results: (a) prior detection map; (b) leakage likelihood map; (c) posterior detection map.

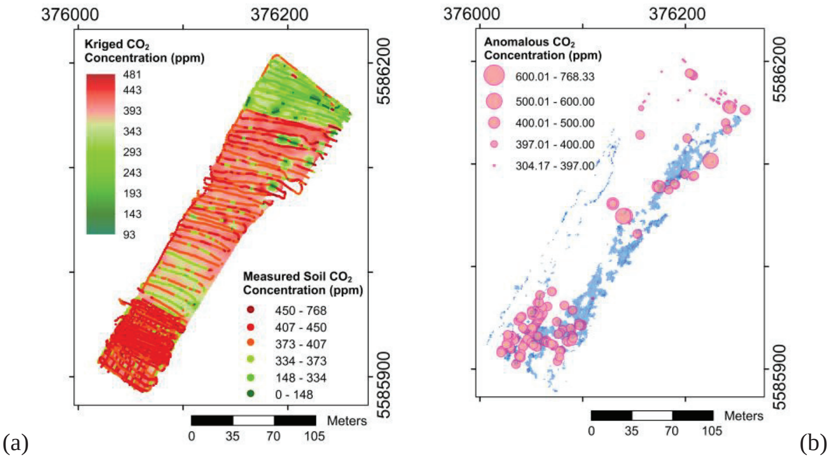


Fig. 4. (a) Mobile laser measurements acquired by BGS superimposed over the kriged CO_2 concentration map; (b) filtered anomalous CO_2 measurements superimposed over the dissonance (posterior) detection map.

measurements were identified (Fig. 4b), thus considering the spatial variation of the background mean CO_2 concentration as the cut-off to determine the anomalous measurements. The result shows a general agreement for the regions containing anomalous values, and demonstrates promise for the methodology.

5. Conclusions

This paper presents a methodology developed to detect surface anomalies corresponding to CO_2 emissions from an onshore natural analogue site using optical remote sensing data. This is an unsupervised approach, based on geostatistical and probabilistic theory, which exploits both the spatial and spectral information contained in multivariate images. The final result is a confidence map showing potential leakage points. The results obtained through the application of the methodology on available airborne hyperspectral datasets for a natural analogue site in Maria Laach, Germany, were validated using mobile laser measurements. These show a good positive correlation with the posterior detection result; however it is acknowledged that more work needs to be done regarding the false or missed detections that have occurred. Future work will also focus on analysing other regions around the Laacher See where ground measurements are available in order to understand the consistency in detection performance of the proposed unsupervised methodology.

Acknowledgements

The authors wish to thank the Instituto Nazionale di Oceanografia e di Geofisica Sperimentale, who provided the hyperspectral data and the British Geological Survey for the ground truth data used in this research. This project was conducted with partial funding by the EU Network of Excellence CO_2GeoNet, Contract No. SES6-2004-CT-502816.

References

[1] Govindan R, Korre A, Durucan S, Imrie CE. A geostatistical and probabilistic spectral image processing methodology for monitoring potential CO_2 leakages on the surface. *International Journal of Greenhouse Gas Control* 2011; **5(3)**: 589-597.
[2] Schminke H-U, Boogaard PVD, Freundt A, Park C. Evolution of complex Plinian eruptions: the late Quaternary Laacher See case history compared with the Minoan eruption. Paper presented at 3rd International Thera Congress, Santorini; 1989.
[3] Aeschbach-Hertig W, Kipfer R, Hofer M, Imboden DM, Wieler R, Signer P. Quantification of gas fluxes from the subcontinental mantle: The example of Laacher See, a maar lake in Germany. *Geochimica et Cosmochimica Acta* 1996. **60(1)**: 31-41.
[4] Giggenbach WF, Sano Y, Schmincke HU. CO_2-rich gases from Lakes Nyos and Monoun, Cameroon; Laacher See, Germany; Dieng, Indonesia, and Mt. Gambier, Australia - variations on a common theme. *Journal of Volcanology and Geothermal Research* 1991. **45(3-4)**: 311-323.
[5] Jones DG, Barlow T, Beaubien SE, Ciotoli G, Lister TR, Lombardi S, May F, Moller I, Pearce JM, Shaw RA. New and established techniques for surface gas monitoring at onshore CO_2 storage sites. *Energy Procedia* 2009. **1(1)**: 2127-2134.
[6] Govindan R, Korre A, Durucan S, Imrie CE. Comparative assessment of the performance of airborne and spaceborne spectral data for monitoring surface CO_2 leakages. *Energy Procedia* 2011. **4**: 3421-3427.
[7] Kitanidis PK. Generalized Covariance Functions in Estimation. *Mathematical Geology* 1993. **25(5)**: 525-540.
[8] Chiles J-P, Delfiner P. *Geostatistics: Modeling Spatial Uncertainty*. New York: John Wiley & Sons; 1999.
[9] Hyvärinen A, Oja E. Independent Component Analysis: algorithms and applications. *Neural Networks* 2000. **13(4-5)**: 411-430.
[10] Kohonen T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 1982. **43(1)**: 59-69.
[11] Dittenbach M, Merkl D, Rauber A. The growing hierarchical self-organizing map. In: *IEEE Proceedings of the International Joint Conference for Neural Networks*, vol. 6; 2000, p. 15-19.
[12] Dempster AP. Upper and Lower Probabilities Induced by a Multivalued Mapping. *The Annals of Mathematical Statistics* 1967. **38(2)**: 325-339.
[13] Shafer G. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press; 1976.